



Accelerating multicolor spectroscopic single-molecule localization microscopy using deep learning

SUNIL KUMAR GAIRE,¹ YANG ZHANG,²  HONGYU LI,¹ RAY YU,²
HAO F. ZHANG,²  AND LESLIE YING^{1,3,*}

¹Department of Electrical Engineering, The State University of New York at Buffalo, Buffalo, NY 14260, USA

²Department of Biomedical Engineering, Northwestern University, Evanston, IL 60208, USA

³Department of Biomedical Engineering, The State University of New York at Buffalo, Buffalo, NY 14260, USA

*leiyang@buffalo.edu

Abstract: Spectroscopic single-molecule localization microscopy (sSMLM) simultaneously provides spatial localization and spectral information of individual single-molecules emission, offering multicolor super-resolution imaging of multiple molecules in a single sample with the nanoscopic resolution. However, this technique is limited by the requirements of acquiring a large number of frames to reconstruct a super-resolution image. In addition, multicolor sSMLM imaging suffers from spectral cross-talk while using multiple dyes with relatively broad spectral bands that produce cross-color contamination. Here, we present a computational strategy to accelerate multicolor sSMLM imaging. Our method uses deep convolution neural networks to reconstruct high-density multicolor super-resolution images from low-density, contaminated multicolor images rendered using sSMLM datasets with much fewer frames, without compromising spatial resolution. High-quality, super-resolution images are reconstructed using up to 8-fold fewer frames than usually needed. Thus, our technique generates multicolor super-resolution images within a much shorter time, without any changes in the existing sSMLM hardware system. Two-color and three-color sSMLM experimental results demonstrate superior reconstructions of tubulin/mitochondria, peroxisome/mitochondria, and tubulin/mitochondria/peroxisome in fixed COS-7 and U2-OS cells with a significant reduction in acquisition time.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Single-molecule localization microscopy (SMLM), including stochastic optical reconstruction microscopy (STORM) [1,2] and photoactivated localization microscopy (PALM) [3,4], have extended the imaging resolution of conventional optical fluorescence microscopy beyond the diffraction limit (~ 250 nm). In these methods, at first random subsets of fluorophores in the sample are imaged in a large number of sequential diffraction-limited frames, then the point spread function (PSF) of detected individual fluorophores in each frame are precisely localized, and finally, all the localization positions from these frames are assembled to generate a super-resolution image. Conventional SMLM provides nanometer-level (~ 20 nm) spatial resolution, but the multicolor function is constrained by spectral cross-talk of fluorescent dyes [5]. Typically, conventional SMLM requires excellent emission spectral separation (~ 100 nm) between dyes to obtain sequential multicolor imaging with minimal cross-talk [5,6]. Recently developed spectroscopic SMLM (sSMLM) simultaneously extracts the spatial locations as well as corresponding spectral information of single-molecule blinking events, offering simultaneous multicolor imaging of multi-stained samples [5,7–13]. In sSMLM, a dispersive optical component, such as a grating or prism, is used to obtain the single-molecule emission spectrum while corresponding spatial

information is collected in a separate optical path [5,8]. Zhang *et al.* [5] employed a dual-objective lens system and a dispersive prism to decouple spatial and spectral information of blinking single molecules and were able to achieve multicolor imaging using dyes having a 10 nm spectral separation. Dong *et al.* [8] used a slit-less monochromator (featuring a blazed diffraction grating) and a mirror to obtain the zero-order (spatial) and the first-order (spectral) images simultaneously enabling multi-label super-resolution imaging from a single round of acquisition. Zhang *et al.* [7] developed a transmission diffraction grating element to obtain the spatial and spectral information of single-molecule blinking events simultaneously and obtain three-color super-resolution images of fixed cells using three dyes with highly overlapping emission spectra. The emission spectra of overlapping dyes can be separated by employing techniques such as spectral regression and customized spectral unmixing algorithms [8,14,15]. Recently a machine learning approach for the robust and accurate spectral classification has also been reported for sSMLM [16].

The ability to acquire and identify distinct spectroscopic signatures from individual single molecules during sSMLM imaging allows simultaneous multicolor super-resolution imaging at sub-diffraction resolution. However, sSMLM typically requires $> 10^4$ sequential diffraction-limited frames to achieve sufficiently dense localizations to reveal details of the biological samples, suggesting long acquisition time and making live-cell and high-throughput imaging challenging. Practically, the acquisition of such long frame sequences also result in dye photobleaching and, consequently, the degradation of image quality. So, a faster sSMLM technique is always desirable. Besides that, sSMLM imaging suffers from cross-color contamination because of intrinsic single-molecule spectral heterogeneity.

Here, we present a novel approach to achieve fast multicolor sSMLM imaging using deep learning. The experimental setup, data acquisition, and spectral classification methods remain the same as previously reported [8,12,13], except that fewer frames (spatial and spectral images) are acquired from multi-dye stained samples, which in turn accelerates the imaging speed. The lower number of frames provides less information on fluorophore localizations and corresponding spectra, not enough for the existing method to extract the fine structures in the sample properly. We employ deep convolution neural networks (CNN) to restore unresolved structures and reconstruct the high-density multicolor super-resolution image using a low-density image without trading off the spatial resolution, which appears to be even superior to those obtained using a large number of frames.

In recent years, deep learning based-approaches had been applied in SMLM. Most of the methods used deep learning to precisely localize the blinking single-molecule PSFs of a large number of frames [17–21], which ultimately accelerate the data processing time of SMLM. A comprehensive review of deep learning methods in SMLM can be found in [22]. For multicolor SMLM imaging, Hershko *et al.* [23] and Kim *et al.* [24] leveraged deep learning for axial localization and color-separation of blinking single-molecules PSFs from a large number of frames. Our method restores the image after performing localization and color-separation (spectral classification) using much fewer frames. The approach is inspired by ANNA-PALM [25], which was developed to accelerate the single-color SMLM imaging using a conditional generative adversarial network (cGANs) [26]. For both training and testing, ANNA-PALM used SMLM and/or widefield images. The novelty of our method includes: First, deep learning is used to accelerate the multicolor sSMLM; Second, single-color SMLM data was used for training and multicolor sSMLM data for testing. Because the training and testing data were acquired with highly different settings, the challenge in our deep learning work is much higher; Finally, we used the residual learning framework [27], a completely different neural network. As a result, our method was able to reduce the cross-color contamination induced by inaccurate spectral classification.

2. Reconstruction method

An experimentally recorded diffraction-limited frame containing a spatial and a spectral image acquired simultaneously is shown in Fig. 1(a). Spatial images were analyzed using standard localization algorithms [28] to determine the location of fluorophore blinking events, and the emission spectra of the corresponding blinking events were recorded from the spectral images. The representative spectra from two individual blinking events highlighted by colored boxes in Fig. 1(a) are shown in Fig. 1(b). Specifically, we obtained a list of localizations $(F_i, x_i, y_i, \lambda_i)_{i=1 \dots n}$ where $F_i \in [1, N]$ is the index of diffraction-limited frames from which localization (x_i, y_i) originates; λ_i is the distinct emission spectra at that location; N is the total frame number; and n is the total number of localizations. The list of localizations can then be separated to multiple imaging channels, according to the pre-defined spectral window of the dyes being used, to visualize the multiple structures in the sample. Finally, the composite multicolor image was obtained by combining the extracted images from all imaging channels. Because localizations from greater than 10^4 frames are often required, the imaging speed of multicolor sSMLM is inherently slow.

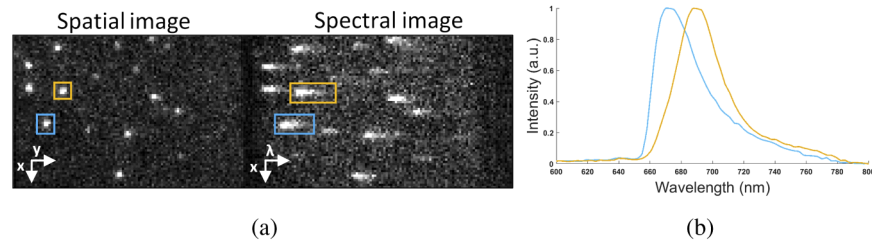


Fig. 1. (a) An experimental sSMLM image frame containing diffraction-limited spatial and spectral images. (b) Emission spectra of two individual blinking events highlighted by the colored box in (a).

Our goal is to reconstruct a high-density multicolor sSMLM image using a low-density multicolor sSMLM image acquired using fewer frames (suppose Q frames and $Q \ll N$). Specifically, after spectral classification, low-density images rendered in each imaging channel, are sparse and incomplete, and we need to restore them. The restoration task can be formulated as an image inpainting task, which aims to restore the missing regions of the corrupted image and reconstruct the original image [29,30]. In recent years, deep learning has been successfully employed for various image restoration problems [25,27,31–37]. Inspired by these successes, we employ a deep learning approach using a deep convolution neural network for restoring the high-density images using the corresponding low-density images acquired in each imaging channel in sSMLM. The deep CNN reconstruction method comprises a training stage and testing stage.

For training, a few single-dye stained super-resolution images for each representative structure of interest (microtubules, mitochondria, and peroxisome) were obtained independently using single-color SMLM. We first acquired N diffraction-limited frames and processed them using standard localization software to obtain high-density super-resolution images. Next, the low-density super-resolution images using much fewer diffraction-limited frames were generated from the same data. The training image set for deep CNN was constructed containing pairs of the low-density images and the corresponding high-density images. Then, the deep CNN was trained using these training images. More detail about the acquisition of training data is explained in section 4.

Once trained, the testing was performed on the new low-density images obtained from the multicolor sSMLM dataset. After accumulating the localization results obtained from a small number of frames (Q frames) with very short acquisition time, $Q\Delta t$, where Δt is the time to acquire a single frame (10 or 20 ms), and applying a pre-defined spectrum cutoff window of

each dye used, low-density images of the specific dye-labeled sample were separated in multiple imaging channels (e.g., the separated structure of mitochondria and tubulin are shown in red and green color, respectively in Fig. 2). The separated low-density single-color images were then fed to their corresponding trained deep CNNs, which rapidly reconstruct the high-density images. Each reconstructed images were then overlaid to get the high-density multicolor super-resolution image.

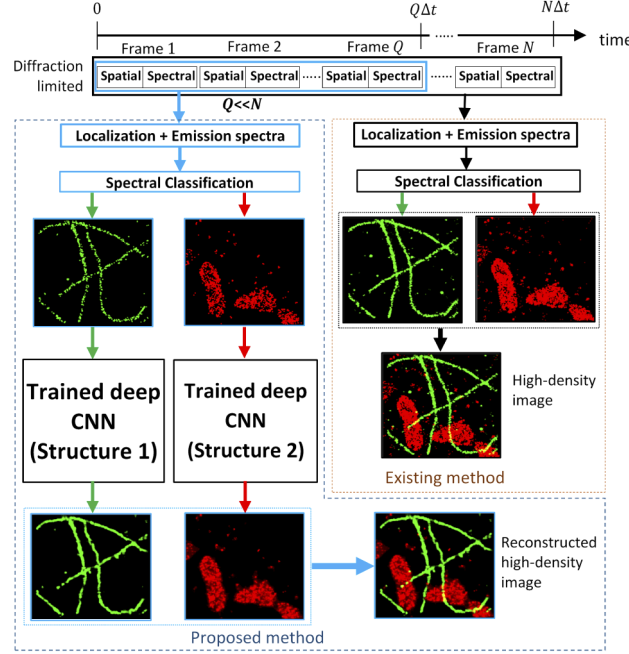


Fig. 2. Comparing our deep learning method and the existing sSMLM method. In existing sSMLM, a large number of frames (suppose N frames) are required to obtain a high-density multicolor super-resolution image. In the proposed method, we use trained deep CNNs to reconstruct the high-density multicolor image using very few frames (suppose Q frames and $Q \ll N$).

3. Deep CNN architecture and learning strategy

Our deep CNN for sSMLM image reconstruction consists of twenty weighted layers ($L = 20$) with a residual learning framework [27,36], as shown in Fig. 3. The input and output of the network are image patches. The dimension of the patches stays constant from the beginning to the end (e.g., 64×64 pixels). For each layer except the first and last one, 64 filters with a kernel size of 3 are used. The first layer operates on the input image. The successive layers learn the feature maps from the input image. The last layer, consisting of a single filter with a kernel size 3, is used for image reconstruction. For all layers except for the last one, the rectified linear units ($ReLU$), $ReLU(x) = \max(0, x)$, is used as an activation function. The output of the last convolution layer is a residual image, which corresponds to the difference between the desired high-density image y and the low-density input image x , and can be written as $r = y - x$. Thus, the final reconstructed image is the sum of network input x and residual image r . The network is trained by minimizing the mean squared error (MSE) which can be written as

$$l(\Theta) = \frac{1}{n} \sum_{i=1}^n \|r_i - f(x_i; \Theta)\|^2, \quad (1)$$

where n is the number of training image (patch) pairs; $\tilde{r}_i = f(x_i; \Theta)$ is the i^{th} network prediction for input x_i , and Θ is the deep learning network parameters to be learned during training. The MSE is a commonly used loss function for training the neural networks where reconstruction accuracy is of key importance. Compared to the conditional generative adversarial network used in ANNA-PALM [25], our network is much simpler and easier to train.

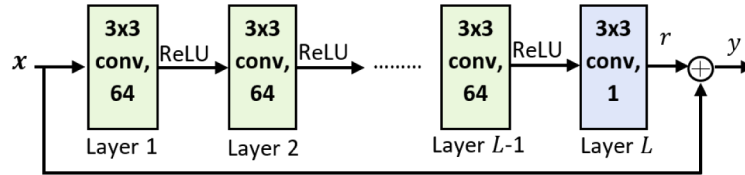


Fig. 3. Residual learning architecture for deep CNN.

The implementation of residual learning tackles the problem of vanishing/exploding gradients [38,39], which hurdle the learning process when the network is very deep. In addition to that, the gradient clipping technique was also used to speed up the training [39,40], which also handles the exploding gradient problems. The network was trained using stochastic gradient descent (SGD) method with Adam optimizer [41] using 30 or more epochs. The learning rate and the threshold value for gradient clipping were set to 10^{-4} and 0.1, respectively, and a batch size of 32 was used. We implemented our model using the TensorFlow framework [42]. Both network training and testing were performed on GTX 1080Ti or GTX 1060 graphical processing units (GPUs) from NVIDIA. Network training takes greater than four hours on a single GPU. Once trained, the high-density reconstruction image is obtained in only ~ 1 second or less.

4. Imaging experimental procedure

4.1. Optical system setup

The sSMLM experimental system has been previously described in detail [7]. In brief, a continuous-wave laser was used for excitation (642 nm, 100 mW, Excelsior one, Spectra-Physics) and guided into an inverted microscope (Ti-E, Nikon), and subsequently focused by a lens (focal length = 400 mm) into the back focal plane of a total internal reflection (TIRF) objective (Nikon CFI Apochromat 100 \times , numerical aperture = 1.49). For conventional SMLM imaging, an EMCCD (iXon 897, Andor) was used to collect image signals. For sSMLM imaging, each frame was split into a 0th and 1st order channels, which respectively provide the unmodified spatial images and spectrally-dispersed first-order spectral images. The spatial and spectral images were then captured by the same EMCCD simultaneously. The illumination power density was ~ 4 kW cm $^{-2}$ at the back focal plane and the exposure time of 10 or 20 ms.

4.2. Cell sample preparation

COS-7 and U2-OS cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) and McCoy's 5A Medium, respectively, supplemented with L-glutamine (2 mM), fetal bovine serum (10% v/v), and penicillin-streptomycin (1% v/v, 100 U mL $^{-1}$) at 37 °C with CO $_2$ (5%). The cells were plated on No.1 borosilicate bottom 8-well Lab-TekTM Chambered Coverglass with 30-50% confluency. The cells were fixed in pre-warmed 3% Paraformaldehyde and 0.1% Glutaraldehyde in Phosphate Buffer Saline (PBS) for 10 min directly, 48 hours after plating. The cells were washed with PBS once, quenched with freshly prepared 0.1% Sodium Borohydride in PBS for 7 min, rinsed with PBS three times at 25 °C and stored at 4 °C.

For SMLM imaging of COS-7 cells, the fixed COS-7 cells were permeabilized with a blocking buffer (3% Bovine Serum Albumin or BSA, 0.2% Triton X-100 in PBS) for 20 min at 25 °C

and then incubated with primary antibodies for tubulin (rabbit anti- β -tubulin, ThermoFisher #PA5-15683, $2.5 \mu\text{g mL}^{-1}$), mitochondria (mouse anti-TOM20, Santa Cruz #sc-17764, $2.5 \mu\text{g mL}^{-1}$), or peroxisome (rabbit anti-PMP70, ThermoFisher #PA1-650, $0.5 \mu\text{g mL}^{-1}$) in blocking buffer overnight at 4°C and rinsed with a washing buffer (0.2% BSA, 0.1% Triton X-100 in PBS) for three times. The cells were further incubated with corresponding donkey secondary antibody ($2.5 \mu\text{g mL}^{-1}$) conjugated with Alexa Fluor 647 (AF647) for 40 min, washed thoroughly with PBS for three times at 25°C and stored at 4°C .

For the multicolor sSMLM imaging experiment, the cell preparation procedure was similar to the single-color imaging method except for the antibodies. For two-color imaging of tubulin and mitochondria, rabbit anti- β -tubulin (ThermoFisher #PA5-15683#, $5 \mu\text{g mL}^{-1}$) and mouse anti-TOM20 (Santa Cruz #sc-17764, $5 \mu\text{g mL}^{-1}$) were used as the primary antibodies, and anti-rabbit AF647 and anti-mouse CF660C ($2.5 \mu\text{g mL}^{-1}$) were added as secondary antibodies. For three-color imaging of tubulin, mitochondria and peroxisome, sheep anti-tubulin (Cytoskeleton #ATN02, $5 \mu\text{g mL}^{-1}$), mouse anti-TOM20 (Santa Cruz #sc-17764, $5 \mu\text{g mL}^{-1}$) and rabbit anti-PMP70, (ThermoFisher #PA1-650, $0.5 \mu\text{g mL}^{-1}$) were used as the primary antibodies, and anti-sheep AF647, anti-mouse CF660C, and anti-rabbit CF680 were used as secondary antibodies.

4.3. Image acquisition

Prior to imaging, an imaging buffer containing 50 mM Tris (pH = 8.0), 10 mM NaCl, 0.5 mg mL^{-1} Glucose Oxidase (Sigma, G2133), 2000 U/mL Catalase (Sigma, C30), 10% (w/v) D-Glucose, and 100 mM Cysteamine were added to each well of the 8-well chambered glass. We recorded 10,000 & 20,000 frames for each SMLM image acquisition with an exposure time of 10 ms as training data. We further collected 20,000, 30,000, and 40,000 frames with an exposure time of 20 ms for the multi-color sSMLM image acquisitions.

5. Data processing, training, and testing setup

Experimental training images with single-dye stained/single-color samples for the tubulin, mitochondria, and peroxisome, were processed separately using established SMLM algorithms. Specifically, the single-molecule blinking events of each diffraction-limited frames were processed using ThunderSTORM [28] plugin of Fiji [43], and localization list results were obtained for each dataset. For tubulin, we used six single-dye stained SMLM datasets with $N = 10,000$ frames. Similarly, six SMLM datasets with $N = 10,000/20,000$ frames were used for the mitochondria and seven SMLM datasets with $N = 10,000$ frames for the peroxisome. These single-color SMLM data can typically be acquired with higher quality than the multi-color sSMLM data (because no spectral classification is needed), and are thereby used for training a neural network that is later used for reconstructing sSMLM images. Although we only used 6/7 SMLM datasets for training each structure, we can generate more than 300,000 inputs (low-density images)/labels (high-density images) pairs to train the corresponding network. This is achieved by (1) randomly selecting $Q = 300 - 1000$ number of frames (for low-density image) out of N ($Q \ll N$) number of frames (for high-density labels) from a single data set and repeat several times, (2) performing data augmentation such as rotating the images, and (3) dividing each image into many overlapping patches of a smaller size (e.g., we used 64×64 here). Thus, only a few SMLM datasets can provide successful training without overfitting. The average shifted histogram method [44] was used as a method of visualization while computationally rendering images in ThunderSTORM using the localization list as an input. The pixel size of all rendered sub-diffraction images was 16 nm. The schematic of training data acquisition and the training process of a deep CNN for a single-dye stained sample of tubulin is shown in Fig. 4. Following a similar procedure, we trained deep CNNs for mitochondria and peroxisome images separately.

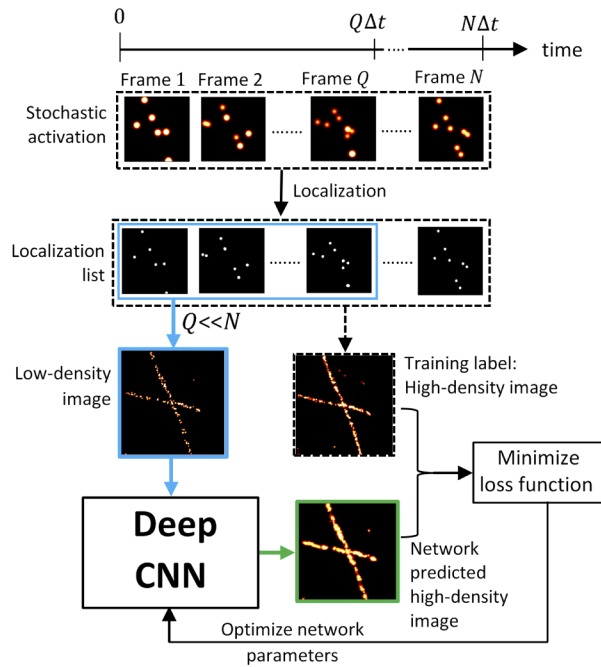


Fig. 4. An overview of SMLM image acquisition for training followed by the training of deep CNN.

For testing, multicolor image data were obtained using sSMLM imaging of completely new samples than that used in training. We used two two-color and two three-color sSMLM imaging datasets for evaluating the performance of our method. The localization lists for each multicolor dataset were obtained by processing the diffraction-limited spatial images using ThunderSTORM. The corresponding emission spectra of every blinking single-molecules were obtained from the spectral image with the spatial image as the reference for the calibration process. The spectral centroid (SC) of each single-molecule was calculated through the weighted average of the wavelength for the measured single-molecule spectrum [5]. The pre-defined spectral windows based on their SCs and spectral precisions were then used to identify and classify the blinking single-molecules from different dyes [7]. Drift correction and necessary density filtering were also performed to each imaging channel data prior to image rendering. Thus, for the testing stage, we imaged very few diffraction-limited frames, performed localization and spectral classification, and then computationally rendered the low-density sSMLM images (of different structures) from each imaging channel (based on the pre-defined spectral windows of the dyes used). These low-density images (e.g., tubulin image from 683-689 nm channel) were then fed to the respective pre-trained deep CNN of that corresponding structure to predict the high-density reconstructed images. After the deep CNN restored each high-density image, all images from the different channels were then overlaid to obtain the reconstructed high-density multicolor super-resolution image. The testing process is illustrated in Fig. 2.

6. Experimental results

6.1. Reconstruction of simultaneous two-color sSMLM imaging

The two-color imaging result of COS-7 cell (Cell 1) with the field-of-view (FOV) of $17.41 \mu\text{m} \times 13.31 \mu\text{m}$ is shown in Fig. 6, where tubulin and mitochondria were labeled with AF647 and CF660C, respectively. The emission spectra of these two dyes are shown in Fig. 5. The

sSMLM two-color localization data for Cell 1 was taken from experiments previously published in [7]. For spectral classification, the spectral window of 683–689 nm (AF647) and 692–698 nm (CF660C) were used. Figure 6(a) shows the composite low-density images obtained from AF647 (tubulin (cyan)) and CF660C (mitochondria (magenta)) channel using 3000 frames with 23,300 localization points. The individual low-density images from tubulin and mitochondrial channels were then fed to the trained deep CNN of that corresponding structure, producing the high-density super-resolution images at the respective CNN output. Overlaying the output images from both networks gives the reconstructed two-color super-resolution image, as shown in Fig. 6(b). Note that all images were normalized, and colors for each channel (different than that of dye) were added to make better visualization. The localization density in the reconstructed image is significantly improved as compared to low-density image and is comparable to (or even slightly better at some positions) the reference image rendered using all 19,997 frames with 134,900 localization points (Fig. 6(c)). The sparse curvilinear structure of tubulins vaguely shown in the low-density image was restored in the reconstruction image with high fidelity, showing the continuous filament structures preserving the sub-diffraction super-resolving capacity of the reference high-density image. Similarly, the mitochondria are denser and restored well. On the other hand, when the localization points are too sparse, the reconstruction using our deep CNN shows some artifacts (indicated using the red arrows in Fig. 6). Such artifacts are primarily due to the lack of information in the low-density input image to recapture the structure. Appendix, Fig. 10 demonstrates that the artifacts can be reduced by increasing the number of frames but at the cost of reduced acceleration.

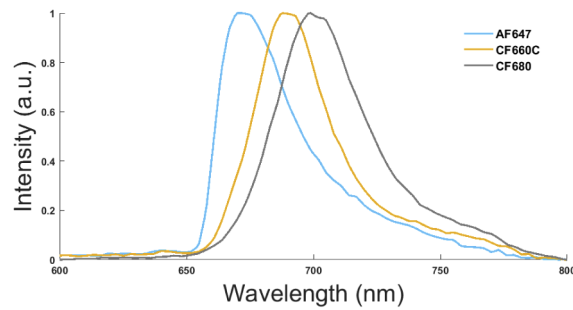


Fig. 5. Normalized emission spectra of AF647, CF660C and CF680 dyes.

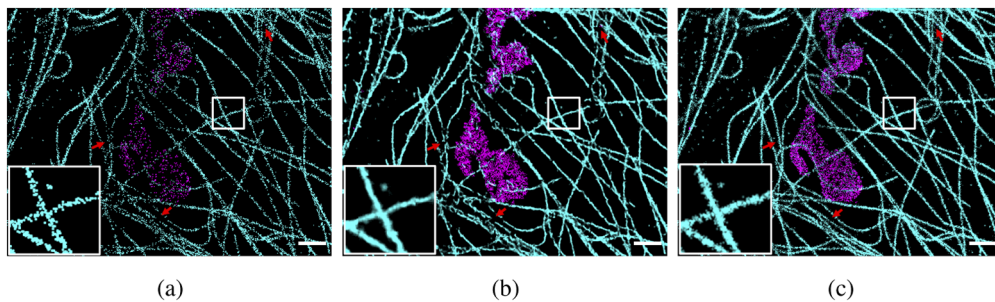


Fig. 6. The two-color sSMLM images of AF647 labeled tubulin (cyan) and CF660C labeled mitochondria (magenta) in COS-7 cell. (a) Low-density image with 3000 frames. (b) Deep CNN reconstruction. (c) High-density image with 19997 frames. Pixel size = 16 nm. Scale bar = 1.5 μ m.

To quantitatively assess the quality of reconstruction, we used a multi-scale structural similarity index (MS-SSIM) [45], a perceptually motivated metric, between the reconstructed and reference image to evaluate the capability of deep CNN in capturing the structural information in the reference image. Since the ground truth is not available for the experimental data, the high-density images with N frames in each spectral channel were used as reference images. The MS-SSIM index has a scale between 0 and 1, with 1 being a perfect match with the reference image. A higher MS-SSIM value indicates a better capture of the structural information. The MS-SSIM values of low-density images of the tubulin and mitochondrial channels of Cell 1 before and after the deep CNN reconstruction are given in Table 1. For both tubulin and mitochondria, MS-SSIM values were improved with the deep CNN reconstruction.

Figure 7 shows the low-density input image (Fig. 7(a)), the reconstructed tubulin image (Fig. 7(b)), and the high-density reference image (Fig. 7(c)) of AF647 (tubulin) channel of Cell 1. The reconstructed image in Fig. 7(b) has much less information from the mitochondria channel compared to the reference image in Fig. 7(c), indicating reduced cross-color contamination. The intensity profiles in Fig. 7(d) also indicate quantitatively, the suppression of contamination in the reconstructed profile. This is an additional advantage that our method offers when reconstructing images using lesser frames. In the existing method, when the frame number increases, the misclassification of spectral signatures in each frame will accumulate to generate substantial cross-color contamination. In contrast, starting with a low-density tubulin image, our trained deep CNN treats the very few misclassifications from mitochondria as “noise” and reduces them because the network is trained to recognize the tubulin structure only. Thus, the CNN reconstructed image from a small number of frames has less cross-color contamination compared to the image obtained from accumulating a large number of frames. However, when the misclassified localizations are so severe that they are visibly present in the low-density input image, the deep CNN fails to remove them completely but produces artifacts (shown by white arrows in Fig. 7(b) and the intensity profile in Fig. 7(d)).

In addition, the full width at half maximum (FWHM) of the intensity profiles across the tubulin filaments is presented in Fig. 7(f). The boxplots were generated from the FWHM measurements taken at ten different positions of the tubulin filaments. Each measurement was performed at the same position for the reconstructed and reference images. Figures 7(h) and 7(i) provide an example of FWHM measurements using the intensity profiles for the orange line segments in Figs. 7(b) and 7(c), respectively. The black dots are measured intensities, and the blue curves are fitted Gaussian functions, with standard deviation σ and FWHM (double black arrow) as indicated. Red lines in Fig. 7(f) show the medians with the values as indicated. The upper and lower ends of the box show the 75th and 25th percentile of the data, respectively. Similarly, whiskers show the full extent of the data. The FWHM was calculated using $\text{FWHM} = 2\sqrt{2\ln 2}\sigma \approx 2.355\sigma$. The FWHM of deep CNN reconstruction is lower than that of the reference image (suggesting better resolution). The FWHM provides only an upper bound of resolution. The measured FWHM based widths are based on the combined size of microtubules labeled with the primary and secondary antibodies (~ 15 nm) and having localization precision of > 15 nm. Thus, the measured widths are compatible with the diameter of the actual microtubules (~ 25 nm). Additionally, Fig. 7(g) shows the intensity profiles along with the red line segments of the images of Figs. 7(a-c). The upper panel is for line segment *a*, and the lower is for line segment *b*. In both cases, the deep CNN is capable of distinguishing and restoring the two nearby filaments by improving the intensities compared to the low-density image. During reconstruction, the deep CNN tends to interpolate the adjacent localization positions, so a slight deviation of the second peak from the reference peak can be observed on the upper panel of Fig. 7(g).

Figure 8 shows the pseudo-colored two-color sSMLM images of immunostained COS-7 cell (Cell 2), with a FOV of $30.72 \mu\text{m} \times 14.34 \mu\text{m}$, with CF660C labeled mitochondria, and CF680 labeled peroxisome. The spectral window of 692-698 nm and 703-709 nm for CF660C and

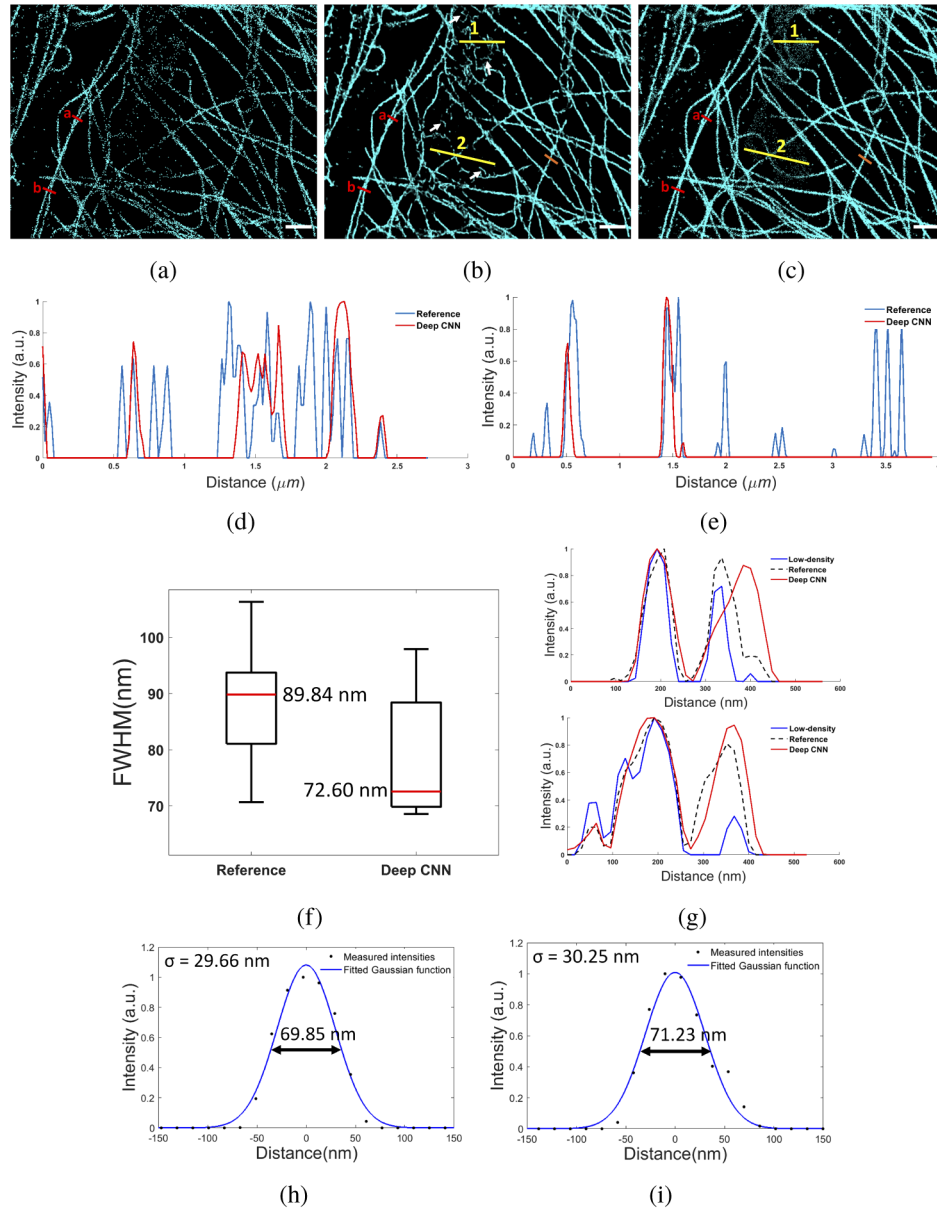


Fig. 7. (a-c) Low-density image, deep CNN reconstruction, and the high-density reference image of the AF647 (tubulin) channel of Cell 1, respectively. Pixel size = 16 nm. Scale bar = 1.5 μm . (d) and (e) The intensity profiles of the yellow line segments 1 and 2 shown in figures (b) and (c). In both cases, it can be observed that the deep CNN reconstruction has less information from the mitochondria channel compared to the reference image, showing reduced cross-color contamination. White arrows in (b) show some artifacts. (f) Boxplots showing the distribution of FWHM measured on ten different positions of the reference tubulin image (c) and deep CNN reconstruction (b). (g) Intensity profiles along with the red line segments shown in figures (a) to (c), with the top for line segment *a* and the bottom for line segment *b*. (h) and (i) Intensity profile and FWHM of the orange line segments shown in figures (b) and (c), respectively. Black dots are measured intensities, and blue curves are fitted Gaussian functions, with standard deviation σ and FWHM (double arrow) as indicated.

CF680, respectively, were used to separate the individual structures into two imaging channels. The overlaid low-density image (Fig. 8(a)) was rendered using 5000 frames with a total of 9,600 localization points. Similarly, the overlaid reference image was rendered using 40,000 frames, with a total of 56,400 localizations. The reconstructed images of mitochondria and peroxisome were obtained by feeding the low-density images obtained after spectral classification to the respectively pre-trained deep CNNs. The overlaid two-color reconstructed image is shown in Fig. 8(b). The MS-SSIM values of the low-density and the reconstructed images are listed in Table 1. The high MS-SSIM values after reconstruction indicate superior reconstruction for both mitochondria and peroxisome with higher structural similarity with the reference images.

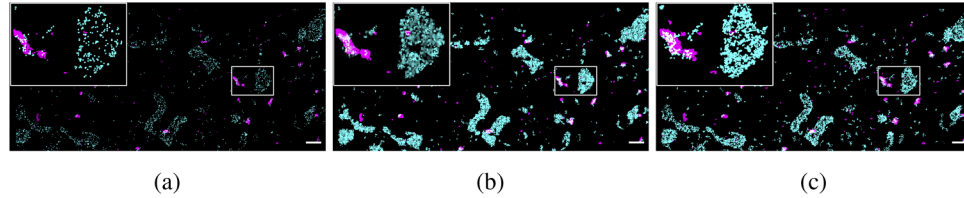


Fig. 8. Pseudo-colored dual-color sSMLM images of CF660C labeled mitochondria (cyan) and CF680 labeled peroxisome (magenta) in immunostained COS-7 cell. (a) Low-density image with 5000 frames. (b) Deep CNN reconstruction. (c) High-density image with 40,000 frames. Pixel size = 16 nm. Scale bar = 1.5 μ m.

Table 1. MS-SSIM values for simultaneous two-color imaging.

COS-7 Cells		Low-density	Deep CNN
Cell 1	Tubulin	0.5413	0.7905
	Mitochondria	0.7593	0.8531
Cell 2	Mitochondria	0.6600	0.9746
	Peroxisome	0.6756	0.8195

6.2. Reconstruction of simultaneous three-color sSMLM imaging

Figure 9 shows pseudo-colored three-color sSMLM images of immunostained U2-OS cells (Cell 3 & Cell 4), with a FOV of $24.58 \mu\text{m} \times 10.24 \mu\text{m}$, with AF647 stained tubulin, CF660C stained mitochondria, and CF680 stained peroxisome. The emission spectra of the three dyes are shown in Fig. 5. The spectral window of 683-689 nm, 692-698 nm, and 703-709 nm were respectively used for the three dyes (AF647, CF660C, and CF680) to separate the individual structures into three imaging channels. Thus, using the spectral signatures obtained from the AF647, CF660C, and CF680 molecular labels, sSMLM resolved the spatial distribution of mitochondria and peroxisome from that of the tubulin. For deep CNN reconstruction, we rendered three low-density images using the localization list obtained from the three imaging channels and used them as the input to each trained deep CNNs of the corresponding structure. For both cells, overlaid low-density images were generated using 4500 frames. Similarly, the overlaid high-density reference images were generated using 30,000 frames. 16,900 and 21,100 localizations were used to generate the low-density images of Figs. 9(a) and 9(d), respectively. Further, 81,800 and 1,20,500 localizations were used to render high-density reference images of Figs. 9(c) and 9(f), respectively.

As shown in the three-color images (Fig. 9) and individual images of each channel of Cell 3 (Appendix, Fig. 11), the deep CNN reconstruction recapitulated most of the features lost in the low-density images. The mitochondria and peroxisome structures are reconstructed well

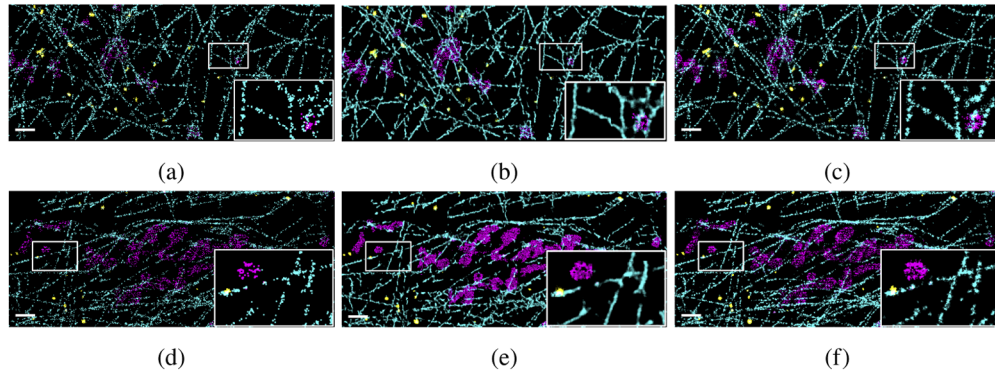


Fig. 9. Pseudo-colored three-color sSMLM images of AF647 labeled tubulin (cyan), CF660C labeled mitochondria (magenta) and CF680 labeled peroxisome (yellow) in immunostained U-2 OS cells. (a-c) Low-density image with 4500 frames, deep CNN reconstruction, and high-density image with 30,000 frames of Cell 3. (d-f) Low-density image with 4500 frames, deep CNN reconstruction, and high-density image with 30,000 frames of Cell 4. Pixel size = 16 nm. Scale bar = 1.5 μ m.

when compared to the reference. On the other hand, the reconstructed tubulin structures show some artifacts. Due to the existing challenges in three-color sSMLM data acquisition, even the high-density reference images of tubulin structures show some discontinuities. The MS-SSIM values of the tubulin, mitochondria, and peroxisome images before and after reconstruction are shown in Table 2. The larger MS-SSIM values of reconstructed images compared to the low-density images indicates a higher similarity with the reference high-density images. It is worth noting that the reference high-density image, obtained from a large number of frames, still might deviate from the ground truth.

Table 2. MS-SSIM values for simultaneous three-color imaging.

U2-OS Cells		Low-density	Deep CNN
Cell 3	Tubulin	0.5214	0.6738
	Mitochondria	0.6742	0.8430
	Peroxisome	0.8446	0.9165
Cell 4	Tubulin	0.5410	0.6571
	Mitochondria	0.5601	0.7441
	Peroxisome	0.8344	0.8861

7. Conclusion

We presented a computational method for fast multicolor sSMLM imaging using deep CNN. Our method reconstructs high-quality multicolor super-resolution images using low-density images rendered from very few diffraction-limited frames, allowing a considerable reduction in sSMLM data acquisition time, without compromising spatial resolution. The experimental results showed superior image reconstruction with a 6.67-fold reduction in the number of frames for simultaneous two-color imaging containing tubulin and mitochondria, and 8-fold reduction for simultaneous imaging of peroxisome and mitochondria in fixed COS-7 cells. Similarly, we also showed improved reconstruction with a 6.67-fold reduction in the number of frames for three-color simultaneous imaging of tubulin, mitochondrial, and peroxisome in fixed U2-OS cells. Additionally, the cross-color contamination, if any, introduced during spectral classification, is

also reduced in the reconstruction images. To accelerate the sSMLM imaging, no change in an existing optical setup, labeling protocol, or spectral classification method is needed. It only requires prior training using high-density images (label) with similar structures.

The proposed method has several limitations. First, we used the single-color SMLM data for training and multicolor sSMLM data for testing. Because the data acquisitions and localization methods are different for single-color and multicolor datasets, the generalizability of the trained network on the testing data might be limited. Second, although our method is able to reduce the cross-color contamination, the contamination may lead to artifacts when the spectral misclassification is severe. Such artifacts might be alleviated by augmenting the training data with the cross-color contaminated cases, which will be investigated in our future studies. Third, when the input image quality is limited due to scarcity of the localization points or increased noise, the reconstructed images may misrepresent the actual structures (broken or extra structures). Such misrepresentation can be alleviated by improving the input image quality using more frames but at the cost of reduced acceleration.

We anticipate that combining spectroscopy, super-resolution optical microscopy, and our deep learning method offers a novel avenue for multicolor, and potentially live-cell and high-throughput imaging to investigate the complex nanoscopic biological structures and their interactions.

Appendix

Figs. 10 & 11, Table 3.

Table 3. MS-SSIM values of Fig. 10 (compared with the reference image of Fig. 6(c)) before and after reconstruction of simultaneous two-color imaging of Cell 1 with various frame numbers.

Frames	Low-density				Deep CNN			
	2500	3500	4500	5500	2500	3500	4500	5500
Tubulin	0.4722	0.6019	0.6929	0.7574	0.7644	0.8009	0.8160	0.8241
Mitochondria	0.7416	0.7753	0.8001	0.8256	0.8587	0.8648	0.8783	0.8936

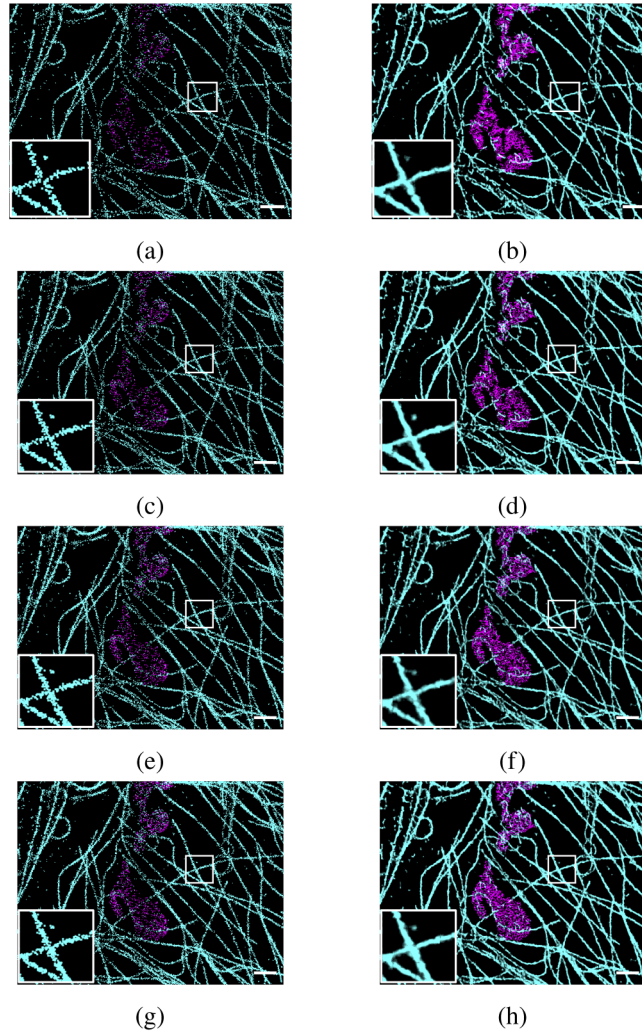


Fig. 10. Deep CNN reconstructions for simultaneous two-color sSMLM images of Cell 1 with various frame numbers. (a, c, e, g) Low-density image with 2500, 3500, 4500, and 5500 frames, respectively. (b, d, f, h) Deep CNN reconstruction of 2500, 3500, 4500 and, 5500 frames images, respectively. The MS-SSIM values are shown in Table 3.

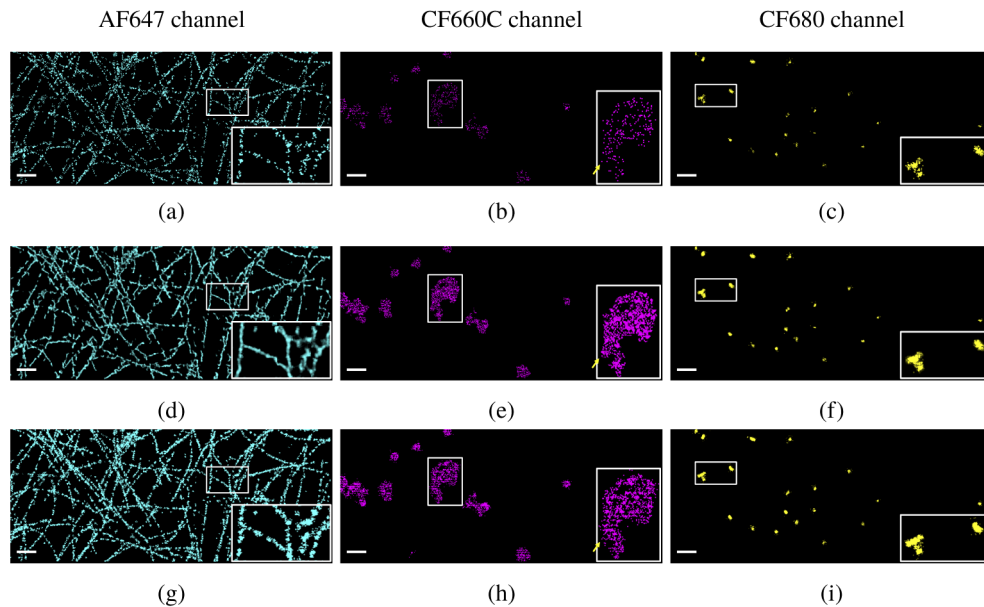


Fig. 11. Deep-CNN reconstruction of three channels of Fig. 9 (Cell 3). (a-c) Low-density images, (d-f) reconstructed images, and (g-i) reference images of each channel, respectively. Pixel size = 16 nm. Scale bar = 1.5 μ m.

Funding

Directorate for Engineering (CBET-1604531, CBET-1706642, EFMA-1830969); National Institutes of Health (R01EY026078, R01EY029121).

Disclosures

The authors declare no conflicts of interest.

References

1. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nat. Methods* **3**(10), 793–796 (2006).
2. S. Van de Linde, A. Löschberger, T. Klein, M. Heidbreder, S. Wolter, M. Heilemann, and M. Sauer, "Direct stochastic optical reconstruction microscopy with standard fluorescent probes," *Nat. Protoc.* **6**(7), 991–1009 (2011).
3. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science* **313**(5793), 1642–1645 (2006).
4. S. T. Hess, T. P. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophys. J.* **91**(11), 4258–4272 (2006).
5. Z. Zhang, S. J. Kenny, M. Hauser, W. Li, and K. Xu, "Ultrahigh-throughput single-molecule spectroscopy and spectrally resolved super-resolution microscopy," *Nat. Methods* **12**(10), 935–938 (2015).
6. G. T. Dempsey, J. C. Vaughan, K. H. Chen, M. Bates, and X. Zhuang, "Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging," *Nat. Methods* **8**(12), 1027–1036 (2011).
7. Y. Zhang, K.-H. Song, B. Dong, J. L. Davis, G. Shao, C. Sun, and H. F. Zhang, "Multicolor super-resolution imaging using spectroscopic single-molecule localization microscopy with optimal spectral dispersion," *Appl. Opt.* **58**(9), 2248–2255 (2019).
8. B. Dong, L. Almassalha, B. E. Urban, T.-Q. Nguyen, S. Khuon, T.-L. Chew, V. Backman, C. Sun, and H. F. Zhang, "Super-resolution spectroscopic microscopy via photon localization," *Nat. Commun.* **7**(1), 12290 (2016).
9. M. J. Mlodzionoski, N. M. Curthoys, M. S. Gunewardene, S. Carter, and S. T. Hess, "Super-resolution imaging of molecular emission spectra and single molecule spectral fluctuations," *PLoS One* **11**(3), e0147506 (2016).
10. K.-H. Song, Y. Zhang, G. Wang, C. Sun, and H. F. Zhang, "Three-dimensional biplane spectroscopic single-molecule localization microscopy," *Optica* **6**(6), 709–715 (2019).

11. M. N. Bongiovanni, J. Godet, M. H. Horrocks, L. Tosatto, A. R. Carr, D. C. Wirthensohn, R. T. Ranasinghe, J.-E. Lee, A. Ponjavic, J. V. Fritz, C. M. Dobson, D. Klenerman, and S. F. Lee, "Multi-dimensional super-resolution imaging enables surface hydrophobicity mapping," *Nat. Commun.* **7**(1), 13544 (2016).
12. K.-H. Song, B. Dong, C. Sun, and H. F. Zhang, "Theoretical analysis of spectral precision in spectroscopic single-molecule localization microscopy," *Rev. Sci. Instrum.* **89**(12), 123703 (2018).
13. B. Dong, J. L. Davis, C. Sun, and H. F. Zhang, "Spectroscopic analysis beyond the diffraction limit," *Int. J. Biochem. Cell Biol.* **101**, 113–117 (2018).
14. T. Zimmermann, "Spectral imaging and linear unmixing in light microscopy," in *Microscopy Techniques*, (Springer, 2005), pp. 245–265.
15. H. Grahn and P. Geladi, *Techniques and Applications of Hyperspectral Image Analysis* (John Wiley & Sons, 2007).
16. Z. Zhang, Y. Zhang, L. Ying, C. Sun, and H. F. Zhang, "Machine-learning based spectral classification for spectroscopic single-molecule localization microscopy," *Opt. Lett.* **44**(23), 5864–5867 (2019).
17. E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," *Optica* **5**(4), 458–464 (2018).
18. N. Boyd, E. Jonas, H. P. Babcock, and B. Recht, "Deeploco: Fast 3D localization microscopy using neural networks," *BioRxiv* p. 267096 (2018).
19. P. Zelger, K. Kaser, B. Rossboth, L. Velas, G. Schütz, and A. Jesacher, "Three-dimensional localization microscopy using deep learning," *Opt. Express* **26**(25), 33166–33179 (2018).
20. A. Speiser, S. C. Turaga, and J. H. Macke, "Teaching deep neural networks to localize sources in super-resolution microscopy by combining simulation-based learning and unsupervised learning," *arXiv preprint arXiv:1907.00770* (2019).
21. E. Nehme, D. Freedman, R. Gordon, B. Ferdman, T. Michaeli, and Y. Shechtman, "Dense three dimensional localization microscopy by deep learning," *arXiv preprint arXiv:1906.09957* (2019).
22. L. Möckl, A. R. Roy, and W. Moerner, "Deep learning in single-molecule microscopy: fundamentals, caveats, and recent developments," *Biomed. Opt. Express* **11**(3), 1633–1661 (2020).
23. E. Hershko, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Multicolor localization microscopy by deep learning," *arXiv preprint arXiv:1807.01637* (2018).
24. T. Kim, S. Moon, and K. Xu, "Information-rich localization microscopy through machine learning," *Nat. Commun.* **10**(1), 1996 (2019).
25. W. Ouyang, A. Aristov, M. Lelek, X. Hoa, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nat. Biotechnol.* **36**(5), 460–468 (2018).
26. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, (2014), pp. 2672–2680.
27. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), pp. 770–778.
28. M. Ovesný, P. Křížek, J. Borkovec, Z. Švindrych, and G. M. Hagen, "ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging," *Bioinformatics* **30**(16), 2389–2390 (2014).
29. Y. Wang, S. Jia, H. F. Zhang, D. Kim, H. Babcock, X. Zhuang, and L. Ying, "Blind sparse inpainting reveals cytoskeletal filaments with sub-nyquist localization," *Optica* **4**(10), 1277–1284 (2017).
30. S. K. Gaire, C. Zhang, H. Li, P. Huang, R. Liu, H. Wang, D. Liang, and L. Ying, "Accelerated 3D localization microscopy using blind sparse inpainting," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, (IEEE, 2019), pp. 526–529.
31. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
32. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), pp. 3431–3440.
33. K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), pp. 449, 458.
34. R. Gao and K. Grauman, "On-demand learning for deep image restoration," in *Proceedings of the IEEE International Conference on Computer Vision*, (2017), pp. 1086–1095.
35. X.-J. Mao, C. Shen, and Y.-B. Yang, "Image denoising using very deep fully convolutional encoder-decoder networks with symmetric skip connections," *arXiv preprint arXiv:1603.09056* **2** (2016).
36. J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), pp. 1646–1654.
37. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica* **4**(11), 1437–1443 (2017).
38. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, (2010), pp. 249–256.
39. R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International conference on machine learning*, (2013), pp. 1310–1318.
40. J. Zhang, T. He, S. Sra, and A. Jadbabaie, "Why gradient clipping accelerates training: A theoretical justification for adaptivity," in *International Conference on Learning Representations*, (2019).

41. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).
42. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv:1603.04467 (2016).
43. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: an open-source platform for biological-image analysis," *Nat. Methods* **9**(7), 676–682 (2012).
44. D. W. Scott, "Averaged shifted histograms: effective nonparametric density estimators in several dimensions," *Ann. Statist.* **13**(3), 1024–1040 (1985).
45. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, vol. 2 (2003), pp. 1398–1402.